

[First Hit](#) [Fwd Refs](#)[Previous Doc](#)[Next Doc](#)[Go to Doc#](#)**End of Result Set**☐ [Generate Collection](#) [Print](#)

L5: Entry 1 of 1

File: USPT

Apr 24, 2001

DOCUMENT-IDENTIFIER: US 6223171 B1

TITLE: What-if index analysis utility for database systems

Detailed Description Text (65):

Database server 220 for one embodiment uses an adaptive page-level sampling technique as described in U.S. patent application Ser. No. 09/139,835 (Attorney Docket No. 14-683) filed on the same date as this patent application, entitled HISTOGRAM CONSTRUCTION USING ADAPTIVE RANDOM SAMPLING WITH CROSS-VALIDATION FOR DATABASE SYSTEMS, by Surajit Chaudhuri, Rajeev Motwani, and Vivek Narasayya, to gather statistical information for the what-if index to be created. This patent application is herein incorporated by reference. Briefly, database server 220 obtains a seed sample of approximately m pages of the table for the what-if index. Database server 220 for one embodiment samples approximately $m=n$ pages of the table where n is the number of pages in the table. Based on this sample, database server 220 generates a sorted list of column values and a set of statistical measures comprising an equi-depth histogram of the column values characterized by step boundaries. Database server 220 samples another approximately m pages of the table and tests how well the values of this new sample fit within the histogram. If this test for convergence fails, database server 220 merges the new sample with the sample (s) collected thus far and updates the sorted list of column values as well as the statistical measures. Database server 220 continues to collect and check new samples in this manner until the values of a new sample fit within the updated histogram within a predetermined degree of accuracy.

Other Reference Publication (4):

Chaudhuri, Surajit, et al., "Random Sampling for Histogram Construction: How Much is Enough?" Proceedings of ACM SIGMOD, Seattle, Washington, pp. 436-447 (Jun. 1-4, 1998).

Other Reference Publication (17):

Olken, Frank, et al., "Simple Random Sampling from Relational Databases," Proceedings of the Twelfth International Conference on Very Large Data Bases (VLDB), Kyoto, pp. 160-169 (Aug. 1986).

Other Reference Publication (18):

Olken, Frank, "Random Sampling from Databases," PhD Dissertation, University of California at Berkeley, Abstract, pp. iii-vi (1993).

Other Reference Publication (19):

Olken, Frank, et al., "Random Sampling from Databases--A Survey," Information and Computing Sciences Div., Lawrence Berkeley Laboratory, Berkeley, California, pp. 1-55 (Mar. 1994).

[Previous Doc](#)[Next Doc](#)[Go to Doc#](#)



US Patent & Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide

(generate or create) and (random sampling) and (parallel data)



THE ACM DIGITAL LIBRARY


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used

generate or create and random sampling and parallel database

Found 81,922 of 148,162

Sort results by

relevance

[Save results to a Binder](#)[Try an Advanced Search](#)

Display results

expanded form

[Search Tips](#)[Try this search in The ACM Guide](#)☐ Open results in a new window

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale ☐ ☐ ☐ ☐ ☐**1 Query evaluation techniques for large databases**

Goetz Graefe

June 1993 **ACM Computing Surveys (CSUR)**, Volume 25 Issue 2

Full text available: pdf(9.37 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

Database management systems will continue to manage large data volumes. Thus, efficient algorithms for accessing and manipulating large sets and sequences will be required to provide acceptable performance. The advent of object-oriented and extensible database systems will not solve this problem. On the contrary, modern data models exacerbate the problem: In order to manipulate large sets of complex objects as efficiently as today's database systems manipulate simple records, query-processi ...

Keywords: complex query evaluation plans, dynamic query evaluation plans, extensible database systems, iterators, object-oriented database systems, operator model of parallelization, parallel algorithms, relational database systems, set-matching algorithms, sort-hash duality

2 Evaluating top-k queries over web-accessible databases

Amélie Marian, Nicolas Bruno, Luis Gravano

June 2004 **ACM Transactions on Database Systems (TODS)**, Volume 29 Issue 2

Full text available: pdf(1.03 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

A query to a web search engine usually consists of a list of keywords, to which the search engine responds with the best or "top" k pages for the query. This top- k query model is prevalent over multimedia collections in general, but also over plain relational data for certain applications. For example, consider a relation with information on available restaurants, including their location, price range for one diner, and overall food rating. A user who queries such a relation might ...

Keywords: Parallel query processing, query optimization, top- k query processing, web databases.

3 Query-based sampling of text databases

Jamie Callan, Margaret Connell

April 2001 **ACM Transactions on Information Systems (TOIS)**, Volume 19 Issue 2

Full text available: pdf(197.24 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

The proliferation of searchable text databases on corporate networks and the Internet causes a database selection problem for many people. Algorithms such as gGLOSS and CORI can automatically select which text databases to search for a given information need, but only if given a set of resource descriptions that accurately represent the contents of each database. The existing techniques for acquiring resource descriptions have significant limitations when used in wide-area networks controlled ...

Keyw rds: distributed information retrieval, query-based sampling, resource ranking, resource selection, server selection

4 Fast detection of communication patterns in distributed executions

Thomas Kunz, Michiel F. H. Seuren

November 1997 **Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research**

Full text available:  pdf(4.21 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...

5 Computing curricula 2001


September 2001 **Journal on Educational Resources in Computing (JERIC)**

Full text available:  pdf(613.63 KB)  html(2.78 KB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)

6 Algorithms for loading parallel grid files

Jianzhong Li, Doron Rotem, Jaideep Srivastava

June 1993 **ACM SIGMOD Record , Proceedings of the 1993 ACM SIGMOD international conference on Management of data**, Volume 22 Issue 2

Full text available:  pdf(823.08 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

The paper describes three fast loading algorithms for grid files on a parallel shared nothing architecture. The algorithms use dynamic programming and sampling to effectively partition the data file among the processors to achieve maximum parallelism in answering range queries. Each processor then constructs in parallel its own portion of the grid file. Analytical results and simulations are given for the three algorithms.

7 Automating parallel simulation using parallel time streams

Victor Yau

April 1999 **ACM Transactions on Modeling and Computer Simulation (TOMACS)**, Volume 9 Issue 2

Full text available:  pdf(194.69 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)


This paper describes a package for parallel steady-state stochastic simulation that was designed to overcome problems caused by long simulation times experienced in our ongoing research in performance evaluation of high-speed and integrated-services communication networks, while maintaining basic statistical rigors of proper analysis of simulation output data. The package, named AKAROA, accepts ordinary (nonparallel) simulation programs, and all further stages of stochastic simulation shou ...

Keywords: distributed simulation, interprocess communication, output analysis methodology, parallel processing, parallel simulation, parallel time streams, spectral analysis, speedup

8 On the development of a site selection optimizer for distributed and parallel database systems

Fotis Barlos, Ophir Frieder


December 1993 **Proceedings of the second international conference on Information and knowledge management**

Full text available:  [pdf\(1.11 MB\)](#) Additional Information: [full citation](#), [references](#), [index terms](#)

9 Industrial sessions: big data: Automating physical database design in a parallel database

Jun Rao, Chun Zhang, Nimrod Megiddo, Guy Lohman

June 2002 **Proceedings of the 2002 ACM SIGMOD international conference on Management of data**


Full text available:  [pdf\(1.38 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Physical database design is important for query performance in a shared-nothing parallel database system, in which data is horizontally partitioned among multiple independent nodes. We seek to automate the process of data partitioning. Given a workload of SQL statements, we seek to determine automatically how to partition the base data across multiple nodes to achieve overall optimal (or close to optimal) performance for that workload. Previous attempts use heuristic rules to make those decision ...

10 Random sampling techniques for space efficient online computation of order statistics of large datasets

Gurmeet Singh Manku, Sridhar Rajagopalan, Bruce G. Lindsay

June 1999 **ACM SIGMOD Record , Proceedings of the 1999 ACM SIGMOD international conference on Management of data**, Volume 28 Issue 2

Full text available:  [pdf\(1.50 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

In a recent paper [MRL98], we had described a general framework for single pass approximate quantile finding algorithms. This framework included several known algorithms as special cases. We had identified a new algorithm, within the framework, which had a significantly smaller requirement for main memory than other known algorithms. In this paper, we address two issues left open in our earlier paper. First, all known and space efficient algorithms for approximate quantile findin ...

11 Join processing in relational databases

Priti Mishra, Margaret H. Eich

March 1992 **ACM Computing Surveys (CSUR)**, Volume 24 Issue 1


Full text available:  [pdf\(4.42 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

The join operation is one of the fundamental relational database query operations. It facilitates the retrieval of information from two different relations based on a Cartesian product of the two relations. The join is one of the most difficult operations to implement efficiently, as no predefined links between relations are required to exist (as they are with network and hierarchical systems). The join is the only relational algebra operation that allows the combining of related tuples fro ...

Keywords: database machines, distributed processing, join, parallel processing, relational algebra

12 Research sessions: selectivity: Hierarchical subspace sampling: a unified framework for high dimensional data reduction, selectivity estimation and nearest neighbor search

Charu C. Aggarwal

June 2002 **Proceedings of the 2002 ACM SIGMOD international conference on Management of data**Full text available:  pdf(1.40 MB)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

With the increased abilities for automated data collection made possible by modern technology, the typical sizes of data collections have continued to grow in recent years. In such cases, it may be desirable to store the data in a reduced format in order to improve the storage, transfer time, and processing requirements on the data. One of the challenges of designing effective data compression techniques is to be able to preserve the ability to use the reduced format directly for a wide range of ...

13 External memory algorithms and data structures: dealing with

massive data

Jeffrey Scott Vitter


June 2001 **ACM Computing Surveys (CSUR)**, Volume 33 Issue 2Full text available:  pdf(828.46 KB)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Data sets in large applications are often too massive to fit completely inside the computers internal memory. The resulting input/output communication (or I/O) between fast internal memory and slower external memory (such as disks) can be a major performance bottleneck. In this article we survey the state of the art in the design and analysis of external memory (or EM) algorithms and data structures, where the goal is to exploit locality in order to reduce the I/O costs. We consider a variety of ...

Keywords: B-tree, I/O, batched, block, disk, dynamic, extendible hashing, external memory, hierarchical memory, multidimensional access methods, multilevel memory, online, out-of-core, secondary storage, sorting

14 On randomization in sequential and distributed algorithms

Rajiv Gupta, Scott A. Smolka, Shaji Bhaskar

March 1994 **ACM Computing Surveys (CSUR)**, Volume 26 Issue 1Full text available:  pdf(8.01 MB)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Probabilistic, or randomized, algorithms are fast becoming as commonplace as conventional deterministic algorithms. This survey presents five techniques that have been widely used in the design of randomized algorithms. These techniques are illustrated using 12 randomized algorithms—both sequential and distributed—that span a wide range of applications, including: primality testing (a classical problem in number theory), interactive probabilistic proofs ...

Keywords: Byzantine agreement, CSP, analysis of algorithms, computational complexity, dining philosophers problem, distributed algorithms, graph isomorphism, hashing, interactive probabilistic proof systems, leader election, message routing, nearest-neighbors problem, perfect hashing, primality testing, probabilistic techniques, randomized or probabilistic algorithms, randomized quicksort, sequential algorithms, transitive tournaments, universal hashing


15 The integration of application and system based metrics in a parallel program performance tool

Jeffrey K. Hollingsworth, R. Bruce Irvin, Barton P. Miller

April 1991 **ACM SIGPLAN Notices, Proceedings of the third ACM SIGPLAN symposium on Principles and practice of parallel programming**, Volume 26 Issue 7

Full text available:  [pdf\(1.21 MB\)](#)Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)**16 On stability and performance of parallel processing systems**

Nicholas Bambos, Jean Walrand

April 1991 **Journal of the ACM (JACM)**, Volume 38 Issue 2Full text available:  [pdf\(1.60 MB\)](#)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

The general problem of parallel (concurrent) processing is investigated from a queuing theoretic point of view. As a basic simple model, consider infinitely many processors that can work simultaneously, and a stream of arriving jobs, each carrying a processing time requirement. Upon arrival, a job is allocated to a processor and starts being executed, unless it is blocked by another one already in the system. Indeed, any job can be randomly blocked by any preceding one, in the se ...

Keywords: database concurrency control, parallel processing, queueing networks, queueing theory, stability theory, subadditive ergodic theory

17 Research sessions: query processing I: A scalable hash ripple join algorithm

Gang Luo, Curt J. Ellmann, Peter J. Haas, Jeffrey F. Naughton

June 2002 **Proceedings of the 2002 ACM SIGMOD international conference on Management of data**Full text available:  [pdf\(1.12 MB\)](#)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Recently, Haas and Hellerstein proposed the hash ripple join algorithm in the context of online aggregation. Although the algorithm rapidly gives a good estimate for many join-aggregate problem instances, the convergence can be slow if the number of tuples that satisfy the join predicate is small or if there are many groups in the output. Furthermore, if memory overflows (for example, because the user allows the algorithm to run to completion for an exact answer), the algorithm degenerates to bl ...

18 The use of regression methodology for the compromise of confidential information in statistical databases

Michael A. Palley, Jeffrey S. Simonoff

November 1987 **ACM Transactions on Database Systems (TODS)**, Volume 12 Issue 4Full text available:  [pdf\(1.39 MB\)](#)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

A regression methodology based technique can be used to compromise confidentiality in a statistical database. This holds true even when the DBMS prevents application of regression methodology to the database. Existing inference controls, including cell restriction, perturbation, and table restriction approaches, are shown to be generally ineffective against this compromise technique. The effect of incomplete supplemental knowledge on the regression methodology based compromise technique is ...

19 LH*—a scalable, distributed data structure

Witold Litwin, Marie-Anna Neimat, Donovan A. Schneider

December 1996 **ACM Transactions on Database Systems (TODS)**, Volume 21 Issue 4Full text available:  [pdf\(780.53 KB\)](#)Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

We present a scalable distributed data structure called LH*. LH* generalizes Linear Hashing (LH) to distributed RAM and disk files. An LH* file can be created from records with primary keys, or objects with OIDs, provided by any number of distributed and autonomous clients. It does not require a central directory, and grows gracefully, through splits of one bucket at a time, to virtually any number of servers. The number of messages per random insertion is


one in general, and three in the w ...

Keywords: algorithms, data structures, distributed access methods, extensible hashing, linear hashing

20 Cactis: a self-adaptive, concurrent implementation of an object-oriented database management system

Scott E. Hudson, Roger King

September 1989 **ACM Transactions on Database Systems (TODS)**, Volume 14 Issue 3

Full text available:  [pdf\(2.65 MB\)](#)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

Cactis is an object-oriented, multiuser DBMS developed at the University of Colorado. The system supports functionally-defined data and uses techniques based on attributed graphs to optimize the maintenance of functionally-defined data. The implementation is self-adaptive in that the physical organization and the update algorithms dynamically change in order to reduce disk access. The system is also concurrent. At any given time there are some number of computations that must be performed t ...

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2004 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)

Welcome to IEEE Xplore®

- ☐ Home
- ☐ What Can I Access?
- ☐ Log-out

Tables of Contents

- ☐ Journals & Magazines
- ☐ Conference Proceedings
- ☐ Standards

Search

- ☐ By Author
- ☐ Basic
- ☐ Advanced
- ☐ CrossRef

Member Services

- ☐ Join IEEE
- ☐ Establish IEEE Web Account
- ☐ Access the IEEE Member Digital Library

IEEE Enterprise

- ☐ Access the IEEE Enterprise File Cabinet

 Print Format

Your search matched **5** of **1105713** documents.
 A maximum of **500** results are displayed, **15** to a page, sorted by **Relevance** in **Descending** order.

Refine This Search:

You may refine your search by editing the current search expression or entering a new one in the text box.

☐ Check to search within this result set
Results Key:

JNL = Journal or Magazine **CNF** = Conference **STD** = Standard

1 Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition

Miller, M.I.; Srivastava, A.; Grenander, U.;

Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] , Volume: 43 , Issue: 11 , Nov. 1995
 Pages:2678 - 2690

[\[Abstract\]](#) [\[PDF Full-Text \(1440 KB\)\]](#) **IEEE JNL**

2 Probabilistic motion planning for parallel mechanisms

Cortes, J.; Simeon, T.;

Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on , Volume: 3 , 14-19 Sept. 2003
 Pages:4354 - 4359 vol.3

[\[Abstract\]](#) [\[PDF Full-Text \(482 KB\)\]](#) **IEEE CNF**

3 The probabilistic method yields deterministic parallel algorithms

Motwani, R.; Naor, J.; Naor, M.;

Foundations of Computer Science, 1989., 30th Annual Symposium on , 30 Oct.-1 Nov. 1989
 Pages:8 - 13

[\[Abstract\]](#) [\[PDF Full-Text \(524 KB\)\]](#) **IEEE CNF**

4 Multiple target direction of arrival tracking

Srivastava, A.; Miller, M.I.; Grenander, U.;

Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] , Volume: 43 , Issue: 5 , May 1995
 Pages:1282 - 1285

[\[Abstract\]](#) [\[PDF Full-Text \(408 KB\)\]](#) **IEEE JNL**

5 Robust estimation of camera translation between two images using a camera with a 3D orientation sensor

Okatani, T.; Deguchi, K.;

Pattern Recognition, 2002. Proceedings. 16th International Conference on , Volume: 1 , 11-15 Aug. 2002

Pages:275 - 278 vol.1

[\[Abstract\]](#) [\[PDF Full-Text \(393 KB\)\]](#) [IEEE CNF](#)

[Home](#) | [Log-out](#) | [Journals](#) | [Conference Proceedings](#) | [Standards](#) | [Search by Author](#) | [Basic Search](#) | [Advanced Search](#) | [Join IEEE](#) | [Web Account](#) | [New this week](#) | [OPAC](#) | [Linking Information](#) | [Your Feedback](#) | [Technical Support](#) | [Email Alerting](#) | [No Robots Please](#) | [Release Notes](#) | [IEEE Online Publications](#) | [Help](#) | [FAQ](#) | [Terms](#) | [Back to Top](#)

Copyright © 2004 IEEE — All rights reserved